9   DATA ANALYSIS

This section outlines general procedures that will be followed to analyze the data collected from the planned investigation. Analysis of the data collected as part of the investigation will be conducted by the Soils Technical Lead. Laboratory results will be evaluated by providing general descriptions of the soil chemistry data and any biological data generated during the investigation. Stations exhibiting exceedances of applicable criteria will be clearly identified. The areas exhibiting such exceedances will be indicated on a map.

Conduction of risk assessment activities typically requires the application of statistics. Often this is restricted to the use of descriptive statistics that describe the data set's centeral tendency and variability. In some cases, comparisons of data sets using inferential statistics can provide the risk assessment with the ability to quantify exposure differences among species, sites, or temporal periods. The risk assessment to be performed for this sampling is based primarily on descriptive statistics to characterize the data being collected; however, depending on the results, data from some of the bioaccumulation and toxicity studies may support inferential statistics. These are discussed in the specific sections of the plan.

To support development of representative concentrations, summary statistics will be computed from analytical laboratory reports. Evaluating the analytical data reports received from the laboratory is often termed as data handling and is the first step in a risk analyses. Data handling procedures follow those recommended by EPA (1990) and ATSDR (2001). For any analytical data set, data qualifiers are assigned to each sample and chemical estimate by the analytical laboratory. Sample data can be qualified for many different reasons, including poor surrogate recovery, blank contamination, or calibration problems. Several qualifiers may be given. In general these are:

R –   Notes that an aspect of the analysis (such as spike recovery) was not within control limits as specified by the sample protocol, therefore, it is recommended that results be rejected from use.

J –   Notes the compound is present but the concentration value is estimated.

B –   For organic data sets, notes the chemical was also detected in the associated analytical lab blank, and thus, the concentrations may reflect some degree of laboratory contamination.

U – Notes the analyte was not present at a concentration able to be identified.

Data flagged with an R are typically discarded from the data set prior to analyses. The J flagged data will be assumed as actual concentrations and used for subsequent analyses based on the values reported.

Data reported with a B flag will be further evaluated. EPA suggests that for chemicals normally available and used in the laboratory, if the concentration detected in the sample is less than ten times the concentration detected in the blank, the analyte concentration should be disregarded because it is likely to be predominantly associated with laboratory practices and not site-specific conditions. If the analyte is not a typical laboratory chemical, then a screen should still be performed, but utilize a more restrictive screen based on five times the concentration noted in the blank.

Data with a U qualifier will be considered as nondetects. Risk assessments typically involve data sets containing values that are lower than limits deemed reliable enough to report as numerical values. These data points are often reported as nondetected and are referred to as censored data sets. The level of censoring is based on the confidence with which the analytical signal can be discerned from the noise. While the concentration may be highly uncertain for some chemicals, it does not necessarily mean that the concentration is zero. Techniques for handling censored data sets can be grouped into three classes: simple substitution, distributional, and robust methods.

Simple substitution methods, the most common method used, involves substituting a single value as a proxy for each value reported as nondetected. Frequently, the values used are zero, the detection limit, or one-half the detection limit. In the worst-case approach, all nondetects are assigned the value of the SQL, which is the lowest level at which a chemical may be accurately quantitated. This approach biases the mean of the data upwards. On the other hand, assigning all nondetects the value of zero biases the mean downwards. The degree to which the results are biased will depend on the number of samples below the detection limit. Hornung and Reed (1990) discuss the merits of assigning a value of one-half the detection limit. Based on a review of similar data sets and the scientific literature, EPA (1992) concludes that for most data sets, simple substitution will be adequate for most exposure assessments. This method will be used within the risk assessment unless the data set appears to be strongly biased. In this case, distributional methods such as the robust method as described by Helsel (1991) will be used.

For a specific chemical in a given exposure area and medium, summary statistics are calculated based on the censored data sets. The statistics reported are as follows:

- **n:** number of samples

- **frequency of detection:** the number of samples reported divided by the total number of samples evaluated

- **minimum detection limit:** the lowest sample specific detection limit reported for nondetected results in the data set being evaluated

- **maximum detection limit:** the maximum sample specific detection limit reported for nondetected results in the data set being evaluated

- **minimum detection:** the minimum concentration detected in the data set being evaluated

- **maximum detection:** the maximum concentration detected in the data set being evaluated

- **mean:** the arithmetic mean concentration of the data set (after censoring is complete)

- **standard deviation:** the standard deviation of the data set (after censoring is complete)

- **upper confidence limit:** the one-tailed 95 percent upper confidence limit for the arithmetic mean. For data sets that are normally distributed, this will be calculated based on a t statistic. For data sets determined to be lognormally distributed, this will be calculated based on an H statistic (EPA, 1993)[1].

---

[1] Per confidence limit of a specific chemical.

This page intentionally left blank.